빅데이터 분석기사

기출문제

From 수제비

#1 다음 중 데이터로부터 의미있는 정보를 추출해 내는 학문으로 통계학과는 달리 정형 또는 비정형을 막론하고 다양한 유형의 데이터를 분석 대상으로 하고 이를 효과적으로 구현하고 전달하는 과정까지 포함한 개념은 무엇인가?

① 데이터 마이닝

② 데이터 사이언스

③ 데이터 알고리즘

④ 데이터 시각화

#2 비정형 데이터에 대한 설명으로 가장 거리가 먼 것은?

① 수집 데이터 각각이 데이터 객체로 구분된다.

② 고정 필드 및 메타데이터(스키마 포함) 가 정의되지 않는다.

③ 데이터 내부의 데이터 구조에 대한 메타 정보가 포함된 구조이다.

④ Crawler, API, RSS 등의 수집 기술을 활용한다.

#3 다음 중 데이터 유형이 다른 것은 무엇인가?

① HTML ② 웹로그

③ 센서데이터 ④ 이미지파일

#4 다음 중 아래에서 설명하는 데이터 처리 기술은 무엇인가?

<설명>

-다양한 형식으로 수집된 데이터를 분석에 용이하도록 일관성 있는 형식으로 만든다.

-평활화, 집계, 일반화, 정규화, 속성 생성 등의 기법이 사용된다.

① 데이터 필터링

② 데이터 변환

③ 데이터 정제

④ 데이터 통합

#5 다음 중 데이터 이상값 발생 원인으로 올바르지 않은 것은?

① 데이터 수집 과정에서 발생할 수 있는 입력 오류

② 데이터를 측정하는 과정에서 발생하는 측정 오류

③ 고의적인 이상값

④ 데이터 입력이 누락된 값

#6 왜도의 값이 0보다 클 경우 평균(Mean)과 최빈값(Mode), 중위수(Mean) 중 가장 작은 값은 무엇인가?

① 최빈값(Mode)

② 중위수(Median)

③ 평균값(Mean)

④ 최빈값과 중위수, 평균값의 크기는 동일하다.

#7 다음 중 표본 통계량이 표본분산일 때 표본분포로 가장 알맞은 것은?

① Z - 분포 ② T - 분포

③ 카이제곱 분포 ④ F - 분포

#8 다음 중 아래에서 설명하는 용어는 무엇인가?

<설명>

의사결정나무에서 하나의 부모 마디로부터 자식 마디들이 형성될 때, 입력변수(Input Variable)의 선택과 범주(Category)의 병합이 이루어질 기준을 의미한다.

① 분류 규칙　　② 통합 기준

③ 정지 규칙　　④ 분리 기준

#9 다음 중 아래 예시의 빈칸( )에 들어가는 활성화 함수는 무엇인가?

< 예시 >

입력층이 직접 출력층에 연결이 되는 단층신경망 (Single Layer Neural Network)에서 활성함수를 (　　) 로 사용하면 로지스틱 회귀 모형과 작동원리가 유사해진다.

① ReLU 함수　　② 시그모이드 함수

③ Softmax함수　　④ 계단(Step)함수

#10 다음 중 서포트 벡터 머신에 대한 설명 중 가장 올바른 것은?

① 분류 및 예측 모두 사용이 가능하다.

② 다른 방법보다 과대 적합의 가능성이 높은 모델이다.

③ 선형으로 분리가 불가능한 분류 문제에는 적용이 불가능하다.

④ 훈련 시간이 상대적으로 빠르고 정확성이 뛰어나다.

#11 다음 중 학습 데이터의 중복을 허용하며 학습 데이터 세트를 나누는 기법이고 복원추출 방법으로 가장 알맞은 것은?

① 배깅　　② 페이스팅

③ 랜덤 서브스페이스　　④ 랜덤 패치

#12 다음 중 주어진 데이터로부터 학습을 통해 모델 내부에서 결정되는 변수로 가장 알맞은 것은?

① 오차　　② 지역 최적점

③ 매개변수　　④ 모멘텀

#13 다음 중 약한 모형(Weak Model)을 순차적으로 적용해 나가는 과정에서 잘 분류된 샘플의 가중치는 낮추고 잘못 분류된 샘플의 가중치는 상대적으로 높여주면서 샘플 분포를 변화시키는 기법으로 가장 알맞은 것은?

① 다수결(Majority Voting)

② 에이다 부스트(AdaBoost)

③ 랜덤 포레스트(Random Forest)

④ 그레디언트 부스트(Gradient Boost)

#14 다음 중 데이터 시각화의 기능으로 옳지 <u>않은</u> 것은?

① 설명 기능　　② 탐색 기능

③ 구현 기능　　④ 표현 기능

#15 반응변수가 범주형인 경우에 적용되는 회귀분석 모형은 무엇인가?

① 단순 회귀 모형　② 다항 회귀 모형
③ 곡선 회귀 모형　④ 로지스틱 회귀모형

#16 다음 중 앙상블 기법의 유형으로 올바르지 <u>않은</u> 것은?

① 배깅　② 부스팅
③ 랜덤 포레스트　④ ReLU

#17 다음이 설명하는 비정형 데이터 분석기법으로 가장 알맞은 것은?

< 설명 >

-테스터 마이닝 기법을 활용하여 웹상의 문서들과 서비스들로부터 정보를 자동적으로 추출, 발견하는 기법이다.

-정보단위인 '노드'와 연결점인 '링크'를 활용한다.

① 텍스트 마이닝(Text Mining)
② 웹 마이닝(Web Mining)
③ 오피니언 마이닝(Opinion Mining)
④ 사회 연결망분석(Social Network Analysis)

#18 명사형으로 변수와 변수의 크기가 순서와 상관없고 단지 이름으로서 의미를 부여할 수 있는 데이터 속성은 무엇인가?

① 명목형　② 순서형
③ 이산형　④ 연속형

#19 다음 중 신뢰할 수 있고, 확장이 용이하며, 분산 컴퓨팅 환경을 지원하는 오픈 소스 소프트웨어는?

① R　② Hadoop
③ HBase　④ Map Reduce

#20 다음 중 비관계형 데이터 저장소로 기존의 전통적인 방식의 RDBMS와 다르게 설계된 DB를 칭하는 용어는?

① RDB　② HDFS
③ NoSQL　④ MySQL